

RAPPORT D'EXAMEN PRELIMINAIRE INTERNATIONAL

(article 36 et règle 70 du PCT)

REC'D 21 DEC 2004

WIPO

PCT

Référence du dossier du déposant ou du mandataire	POUR SUITE A DONNER voir la notification de transmission du rapport d'examen préliminaire international (formulaire PCT/PEA/416)	
Demande internationale No. PCT/FR 03/02770	Date du dépôt international (<i>jour/mois/année</i>) 19.09.2003	Date de priorité (<i>jour/mois/année</i>) 24.09.2002
Classification internationale des brevets (CIB) ou à la fois classification nationale et CIB G10L15/18		
Déposant THALES		

1. Le présent rapport d'examen préliminaire international, établi par l'administration chargée de l'examen préliminaire international, est transmis au déposant conformément à l'article 36.

2. Ce RAPPORT comprend 5 feuilles, y compris la présente feuille de couverture.

- Il est accompagné d'ANNEXES, c'est-à-dire de feuilles de la description, des revendications ou des dessins qui ont été modifiées et qui servent de base au présent rapport ou de feuilles contenant des rectifications faites auprès de l'administration chargée de l'examen préliminaire international (voir la règle 70.16 et l'instruction 607 des Instructions administratives du PCT).

Ces annexes comprennent 9 feuilles.

3. Le présent rapport contient des indications et les pages correspondantes relatives aux points suivants :

- I Base de l'opinion
- II Priorité
- III Absence de formulation d'opinion quant à la nouveauté, l'activité inventive et la possibilité d'application industrielle
- IV Absence d'unité de l'invention
- V Déclaration motivée selon la règle 66.2(a)(ii) quant à la nouveauté, l'activité inventive et la possibilité d'application industrielle; citations et explications à l'appui de cette déclaration
- VI Certains documents cités
- VII Irrégularités dans la demande internationale
- VIII Observations relatives à la demande internationale

Date de présentation de la demande d'examen préliminaire internationale 13.04.2004	Date d'achèvement du présent rapport 21.12.2004
Nom et adresse postale de l'administration chargée de l'examen préliminaire international Office européen des brevets - P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk - Pays Bas Tél. +31 70 340 - 2040 Tx: 31 651 epo nl Fax: +31 70 340 - 3016	Fonctionnaire autorisé Quélavoine, R N° de téléphone +31 70 340-3946



RAPPORT D'EXAMEN PRÉLIMINAIRE INTERNATIONAL

Demande internationale n°

PCT/FR 03/02770

I. Base du rapport

1. En ce qui concerne les éléments de la demande internationale (*les feuilles de remplacement qui ont été remises à l'office récepteur en réponse à une invitation faite conformément à l'article 14 sont considérées, dans le présent rapport, comme "initiallement déposées" et ne sont pas jointes en annexe au rapport puisqu'elles ne contiennent pas de modifications (règles 70.16 et 70.17)*) :

Description, Pages

1-7

reçue(s) le 23.09.2004 avec lettre du 17.09.2004

Revendications, No.

1-9

recue(s) le 23.09.2004 avec lettre du 17.09.2004

Dessins, Feuilles

1/3-3/3

telles qu'initialement déposées

2. En ce qui concerne la **langue**, tous les éléments indiqués ci-dessus étaient à la disposition de l'administration ou lui ont été remis dans la langue dans laquelle la demande internationale a été déposée, sauf indication contraire donnée sous ce point.

Ces éléments étaient à la disposition de l'administration ou lui ont été remis dans la langue suivante: , qui est:

- la langue d'une traduction remise aux fins de la recherche internationale (selon la règle 23.1(b)).
 - la langue de publication de la demande internationale (selon la règle 48.3(b)).
 - la langue de la traduction remise aux fins de l'examen préliminaire internationale (selon la règle 55.2 ou 55.3).

3. En ce qui concerne les séquences de nucléotides ou d'acide aminé divulguées dans la demande internationale (le cas échéant), l'examen préliminaire internationale a été effectué sur la base du listage des séquences :

- contenu dans la demande internationale, sous forme écrite.
 - déposé avec la demande internationale, sous forme déchiffrable par ordinateur.
 - remis ultérieurement à l'administration, sous forme écrite.
 - remis ultérieurement à l'administration, sous forme déchiffrable par ordinateur.
 - La déclaration, selon laquelle le listage des séquences par écrit et fourni ultérieurement ne va pas au-delà de la divulgation faite dans la demande telle que déposée, a été fournie.
 - La déclaration, selon laquelle les informations enregistrées sous déchiffrable par ordinateur sont identiques à celles du listages des séquences Présenté par écrit, a été fournie.

- #### 4. Les modifications ont entraîné l'annulation :

- de la description, pages :
 - des revendications, nos :
 - des dessins, feuilles :

5. Le présent rapport a été formulé abstraction faite (de certaines) des modifications, qui ont été considérées comme allant au-delà de l'exposé de l'invention tel qu'il a été déposé, comme il est indiqué ci-après (règle 70.2(c)) :

(Toute feuille de remplacement comportant des modifications de cette nature doit être indiquée au point 1 et annexée au présent rapport.)

6. Observations complémentaires, le cas échéant :

V. Déclaration motivée selon l'article 35(2) quant à la nouveauté, l'activité inventive et la possibilité d'application industrielle; citations et explications à l'appui de cette déclaration

1. Déclaration Nouveauté	Oui:	Revendications	4-6,8-9
	Non:	Revendications	1-3,7
Activité inventive	Oui:	Revendications	
	Non:	Revendications	1-9
Possibilité d'application industrielle	Oui:	Revendications	1-9
	Non:	Revendications	

2. Citations et explications

voir feuille séparée

Concernant le point V

V.1 Il est fait référence aux documents suivants :

- D1: US-A-6 141 661 (BROWN ET AL) 31 octobre 2000 (2000-10-31)
D2: US 2002/035471 A1 (BRETON) 21 mars 2002 (2002-03-21)

V.2 La présente demande ne remplit pas les conditions énoncées dans l'article 33(1) PCT, l'objet des revendications 1, 7 n'étant pas conforme au critère de nouveauté défini par l'article 33(2) PCT.

V.2.1 Le document D1 décrit (abrégé, fig.2,3, col.1 l.50-56, col.3 l.22-26, col.3 l.55-65) un système de reconnaissance de la parole dans lequel après un erreur de reconnaissance, la grammaire utilisée pour le décodage est modifiée de façon à inhiber la solution erronée. Le document D1 mentionne la reconnaissance d'identificateurs qui peuvent être constitués de n'importe quelle combinaison de caractères alphanumériques. De fait, certaines de ces combinaisons ne pourront être lues comme un mot mais que comme une phrase.

Par exemple, si l'identificateur à reconnaître est AE439J, il sera interprété par le système de reconnaissance comme le résultat des phrases: [A E 4 3 9 J] ou [A E 439 J] ou encore [A E 43 9 J]...

Si lors d'une première reconnaissance, l'identificateur AA439J est par exemple reconnu, la grammaire de reconnaissance sera modifiée suite à l'action de l'utilisateur pour exclure cette solution lors d'une seconde passe.

L'objet des revendications 1 et 7 est donc connu de D1.

V.2.2 Les revendications dépendantes 2-3 ne contiennent aucune caractéristique qui, en combinaison avec celles de l'une quelconque des revendications à laquelle elles se réfèrent, définisse un objet qui satisfasse aux exigences du PCT en ce qui concerne la nouveauté, les caractéristiques correspondantes sont elles aussi connues de D1.

V.2.3 Les revendications dépendantes 4-6, 8, 9 ne contiennent aucune caractéristique qui, en combinaison avec celles de l'une quelconque des revendications à laquelle elles se réfèrent, définisse un objet qui satisfasse aux exigences du PCT en ce qui concerne l'activité inventive.

À propos des revendications 8 et 9, les caractéristiques sont connues de D2 (abrégé, par.13), et leur intégration dans le système défini dans D1 est une option d'implémentation connue de l'homme de l'art et ne nécessitant aucune activité inventive.

À propos des revendications 4-6, le calcul anticipé de solutions alternatives est standard dans le domaine (choix de la deuxième solution la plus probable, principe du N-Best), son implémentation ne requiert là encore aucune activité inventive de la part de l'homme de l'art.

Procédé de reconnaissance vocale avec correction automatique¹⁰⁰

La présente invention concerne un procédé de reconnaissance vocale avec correction automatique dans les systèmes de reconnaissance vocale à syntaxe contrainte, c'est-à-dire que les phrases reconnaissables se trouvent dans un ensemble de possibilités déterminées. Ce procédé est particulièrement adapté à la reconnaissance vocale en milieu bruité, par exemple dans les cockpits d'avions d'arme ou civil, dans les hélicoptères ou dans l'automobile.

De nombreux travaux dans le domaine de la reconnaissance vocale à syntaxe contrainte ont permis d'obtenir des taux de reconnaissance de l'ordre de 95%, et ce, même dans l'environnement bruité d'un cockpit d'avion d'arme (environ 100-110 dBA autour du casque du pilote). Cependant, cette performance n'est pas suffisante pour faire de la commande vocale un média de commande primaire pour des paramètres critiques du point de vue de la sécurité de vol.

Une stratégie utilisée consiste à soumettre les commandes critiques à une validation du pilote, qui vérifie par la phrase reconnue, que les bonnes valeurs vont être affectées aux bons paramètres (« feedback primaire »). En cas d'erreur du système de reconnaissance – ou erreur de prononciation du pilote – le pilote doit énoncer à nouveau toute la phrase, et la probabilité d'erreur sur la reconnaissance de la phrase à nouveau prononcée est la même. Ainsi par exemple, si le pilote énonce « Select altitude two five five zero feet », le système effectue les algorithmes de reconnaissance et donne un retour visuel au pilote. En envisageant le cas où une erreur se produit, le système va par exemple proposer « SEL ALT 2 5 9 0 FT ». Dans un système classique, le pilote doit alors prononcer de nouveau toute la phrase, avec les mêmes probabilités d'erreur.

Un système de correction d'erreur meilleur en terme de taux de reconnaissance consiste à faire prononcer au pilote une phrase de correction qui sera reconnue comme telle. Par exemple, si l'on reprend l'exemple précédent, le pilote pourra prononcer « Correction third digit five ». Cependant cette méthode augmente la charge de travail du pilote dans le procédé de reconnaissance, ce qui n'est pas souhaitable.

Il est connu de l'art antérieur, voir par exemple US-A-6 141 661, un procédé de reconnaissance vocale d'un identifiant parmi un ensemble

d'identifiants pré-enregistré, dans lequel si un premier identifiant a été reconnu puis invalidé par l'utilisateur, la reconnaissance vocale est réitérée en supprimant le premier identifiant dudit ensemble. Ce procédé cependant ne peut s'appliquer à la reconnaissance vocale de phrases, qui forment un 5 nombre de combinaisons trop important pour être pré-enregistrées.

L'invention propose un procédé de reconnaissance vocale qui met en œuvre une correction automatique de la phrase prononcée permettant d'obtenir un taux de reconnaissance proche de 100%, sans augmentation de la charge du pilote.

10 Pour cela, l'invention concerne un procédé de reconnaissance vocale d'un signal de parole prononcé par un locuteur avec correction automatique, comprenant notamment une étape de traitement dudit signal de parole délivrant un signal sous une forme compressée, une étape de reconnaissance de formes pour rechercher, à partir d'une syntaxe formée 15 d'un ensemble de phrases qui représentent l'ensemble des chemins possibles entre un ensemble de mots préenregistré lors d'une phase préalable, une phrase de ladite syntaxe la plus proche dudit signal sous sa forme compressée, et caractérisé en ce qu'il comprend

- la mémorisation (16) du signal sous sa forme compressée,
- la génération (17) d'une nouvelle syntaxe (SYNT2) dans laquelle le chemin correspondant à ladite phrase déterminée lors de l'étape de reconnaissance antérieure est interdit,
- la réitération de l'étape de reconnaissance de formes pour rechercher, à partir de la nouvelle syntaxe, une autre phrase la plus proche dudit signal mémorisé.

25 D'autres avantages et caractéristiques apparaîtront plus clairement à la lecture de la description qui suit, illustrée par les figures annexées qui représentent :

- la figure 1, le schéma de principe d'un système de reconnaissance vocale de type connu;
- la figure 2, le schéma d'un système de reconnaissance vocale du type de celui de la figure 1 mettant en œuvre le procédé selon l'invention ;
- la figure 3, un schéma illustrant la modification de la syntaxe dans le procédé selon l'invention.

Sur ces figures, les éléments identiques sont référencés par les mêmes repères.

La figure 1 présente le schéma de principe d'un système de reconnaissance vocale à syntaxe contrainte de type connu, par exemple un système embarqué dans un environnement fortement bruité. Dans un système à syntaxe contrainte mono locuteur, une phase d'apprentissage hors temps réel permet à un locuteur donné d'enregistrer un ensemble de références acoustiques (mots) stockés dans un espace de références 10. La syntaxe 11 est formée d'un ensemble de phrases qui représentent l'ensemble des chemins ou transitions possibles entre les différents mots. Typiquement, quelques 300 mots sont enregistrés dans l'espace de référence qui forment typiquement 400 000 phrases possibles de la syntaxe.

Classiquement, un système de reconnaissance vocale comporte au moins trois blocs comme illustré sur la figure 1. Il comporte un bloc 12 d'acquisition du signal de parole (ou prise de son), un bloc 13 de traitement du signal et un bloc 14 de reconnaissance de formes. Une description détaillée de l'ensemble de ces blocs selon un mode de réalisation se trouve par exemple dans la demande de brevet français FR 2 808 917 au nom de la déposante.

De façon connue, le signal acoustique traité par le bloc de prise de son 12 est un signal de parole capté par un transducteur électroacoustique. Ce signal est numérisé par échantillonnage et découpé en un certain nombre de trames recouvrantes ou non, de même durée ou non. Dans le bloc 13 de traitement du signal, on associe classiquement chaque trame à un vecteur de paramètres qui traduit l'information acoustique contenue dans la trame. Il y a plusieurs méthodes pour déterminer un vecteur de paramètres. Un exemple classique de méthode est celle qui utilise les coefficients cepstraux de type MFCC (abréviation de l'expression anglo-saxonne « Mel Frequency Cepstral Coefficient »). Le bloc 13 permet de déterminer dans un premier temps l'énergie spectrale de chaque trame dans un certain nombre de canaux fréquentiels ou fenêtres. Il délivre pour chacune des trames une valeur d'énergie spectrale ou coefficient spectral par canal fréquentiel. Il effectue ensuite une compression des coefficients spectraux obtenus pour tenir compte du comportement du système auditif humain. Il effectue enfin une transformation des coefficients spectraux compressés, ces coefficients

spectraux compressés transformés sont les paramètres du vecteur de paramètres recherché.

Le bloc 14 de reconnaissance de formes est relié à l'espace de références 10. Il compare la série des vecteurs de paramètres issue du bloc 5 de traitement du signal aux références obtenues lors de la phase d'apprentissage, ces références traduisant les empreintes acoustiques de chaque mot, chaque phonème, plus généralement de chaque commande et que l'on appellera de façon générique « phrase » dans la suite de la 10 description. Puisque la reconnaissance de formes s'effectue par comparaison entre vecteurs de paramètres, on doit avoir à disposition ces vecteurs de paramètres de base. On les obtient de la même manière que pour les trames de signal utile, en calculant pour chaque trame de base son énergie spectrale dans un certain nombre de canaux fréquentiels et en utilisant des fenêtres de pondération identiques.

15 A l'issue de la dernière trame, ce qui correspond généralement à la fin d'une commande, la comparaison donne soit une distance entre la commande testée et des commandes de référence, la commande de référence présentant la distance la plus faible est reconnue, soit une probabilité pour que la série des vecteurs de paramètres appartiennent à une 20 suite de phonèmes. Les algorithmes classiquement utilisés pendant la phase de reconnaissance de formes sont dans le premier cas de type DTW (abréviation de l'expression anglo-saxonne pour Dynamic Time Warping) ou, dans le second cas de type HMM (abréviation de l'expression anglo-saxonne Hidden Markov Models). Dans le cas d'un algorithme de type HMM, les 25 références sont des fonctions gaussiennes associées chacune à un phonème et non à des séries de vecteurs de paramètres. Ces fonctions gaussiennes sont caractérisées par leur centre et leur écart-type. Ce centre et cet écart type dépendent des paramètres de toutes les trames du phonème, c'est à dire des coefficients spectraux compressés de toutes les 30 trames du phonème.

Les signaux numériques représentant une phase reconnue sont transmis à un dispositif 15 qui réalise le couplage avec l'environnement, par exemple par affichage de la phrase reconnue sur le viseur tête haute d'un cockpit d'avion.

Comme cela a été précédemment expliqué, pour les commandes critiques, le pilote peut avoir à sa disposition un bouton de validation permettant l'exécution de la commande. Dans le cas où la phrase reconnue serait erronée, il doit généralement répéter la phrase avec une probabilité identique d'erreur.

Le procédé selon l'invention permet une correction automatique de grande efficacité et simple à mettre en œuvre. Son implantation dans un système de reconnaissance vocale du type de la figure 1 est schématisée sur la figure 2.

10 Selon l'invention, à l'issu de la phase de traitement du signal 13, on mémorise (étape 16) le signal de parole sous sa forme compressée (ensemble des vecteurs de paramètres également appelés « cepstres »). Dès qu'une phrase est reconnue, on génère une nouvelle syntaxe (étape 17) dans laquelle la phrase reconnue n'est plus un chemin possible de la 15 syntaxe. On réitère alors la phase de reconnaissance de formes avec le signal mémorisé mais sur la nouvelle syntaxe. Préférentiellement, la reconnaissance de formes est réitérée de manière systématique pour préparer une autre solution possible. Si le pilote détecte une erreur dans la 20 commande reconnue, il appuie par exemple sur un bouton spécifique de correction, ou exerce un appui court ou un double clic sur l'alternat de commande vocale et le système lui propose la nouvelle solution trouvée lors de la réitération de la reconnaissance de formes. On réitère les étapes 25 précédentes pour générer de nouvelles syntaxes qui interdisent toutes les solutions précédemment trouvées. Quand le pilote voit la solution qui correspond réellement à la phrase énoncée, il valide par un moyen quelconque (bouton, voix, etc.).

Reprendons l'exemple cité précédemment en tirant bénéfice de l'invention. Le pilote énonce selon cet exemple « Select altitude two five five zero feet ». Le système effectue les algorithmes de reconnaissance et, par 30 exemple à cause du bruit ambiant, reconnaît « Select altitude two five nine zero feet ». Un feedback visuel est donné au pilote : « SEL ALT 2 5 9 0 FT ». Alors que le locuteur est en train de lire la phrase reconnue, le système 35 anticipe une éventuelle erreur en générant de façon automatique une nouvelle syntaxe dans laquelle la phrase reconnue est supprimée et en réitérant l'étape de reconnaissance de formes.

La figure 3 illustre par un schéma simple, dans le cas de l'exemple précédent, la modification de la syntaxe permettant avec un algorithme de reconnaissance de formes de type DTW la recherche d'une nouvelle phrase.

La phrase énoncée par le locuteur selon l'exemple précédente est « SEL

- 5 ALT 2 5 5 0 FT ». Nous supposons que la phrase reconnue par la première phase de reconnaissance de formes est « SEL ALT 2 5 9 0 FT ». Cette première phase fait appelle à la syntaxe d'origine SYNT1, dans laquelle toutes les combinaisons (ou chemins) sont possibles pour les quatre chiffres à reconnaître. Lors d'une deuxième phase de reconnaissance de formes, la 10 phrase reconnue est écartée des combinaisons possibles, modifiant ainsi l'arbre syntaxique comme cela est illustré sur la figure 3. Une nouvelle syntaxe est générée qui interdit le chemin correspondant à la solution reconnue. Une deuxième phase est alors reconnue. La phase de reconnaissance de formes peut être réitérée avec, à chaque fois, génération 15 d'une nouvelle syntaxe qui reprend la syntaxe précédente mais dans laquelle est supprimée la phrase précédemment trouvée.

Ainsi, la nouvelle syntaxe est obtenue par réorganisation de la syntaxe antérieure de telle sorte à particulariser le chemin correspondant à la phrase déterminée lors de l'étape de reconnaissance antérieure, puis en 20 éliminant ce chemin. Cette réorganisation est faite par exemple en parcourant la syntaxe antérieure en fonction des mots de la phrase préablement reconnue et en formant au fil de ce parcours le chemin spécifique à cette phrase.

Dans un mode de fonctionnement possible, le pilote indique au 25 système qu'il désire une correction (par exemple par un appui court de l'alternat commande vocale) et dès qu'une nouvelle solution est disponible, elle est affichée. La recherche automatique d'une nouvelle phrase s'arrête par exemple lorsqu'une phrase reconnue est validée par le pilote. Dans notre exemple, il est probable que dès la deuxième phase de reconnaissance de 30 formes, le pilote voit « SEL ALT 2 5 5 0 FT ». Il peut alors valider la commande. Dans la mesure où de nombreuses erreurs de reconnaissance sont dues à des confusions entre des mots proches (par exemple, five-nine), l'invention permet de corriger presque à coup sûr ces erreurs avec un minimum de charge de travail supplémentaire pour le pilote et de façon très

rapide du fait de l'anticipation sur la correction que peut effectuer le procédé selon l'invention.

En outre, en générant une nouvelle syntaxe et en réitérant l'étape de reconnaissance de formes sur la nouvelle syntaxe, on n'accroît pas la complexité de l'arbre syntaxique. L'algorithme de traitement peut donc effectuer la reconnaissance avec un délai similaire à chaque itération, ce délai étant imperceptible pour le pilote du fait de l'anticipation de la correction.

10

EPO - DG

23.09.2004

REVENDICATIONS

(100)

- 5 1- Procédé de reconnaissance vocale d'un signal de parole prononcé par un locuteur avec correction automatique, comprenant notamment une étape (13) de traitement dudit signal de parole délivrant un signal sous une forme compressée, une étape (14) de reconnaissance de formes pour rechercher, à partir d'une syntaxe (SYNT1) formée d'un ensemble de phrases qui représentent l'ensemble des chemins possibles entre un ensemble de mots préenregistré lors d'une phase préalable, une phrase de ladite syntaxe la plus proche dudit signal sous sa forme compressée, et caractérisé en ce qu'il comprend
- 10 - la mémorisation (16) du signal sous sa forme compressée,
- 15 - la génération (17) d'une nouvelle syntaxe (SYNT2) dans laquelle le chemin correspondant à ladite phrase déterminée lors de l'étape de reconnaissance antérieure est interdit,
- 20 - la réitération de l'étape de reconnaissance de formes pour rechercher, à partir de la nouvelle syntaxe, une autre phrase la plus proche dudit signal mémorisé.
- 25 2- Procédé de reconnaissance vocale selon la revendication 1, dans lequel la nouvelle syntaxe est obtenue par réorganisation de la syntaxe antérieure de telle sorte à particulariser ledit chemin correspondant à la phrase déterminée lors de l'étape de reconnaissance antérieure, puis élimination de ce chemin.
- 30 3- Procédé de reconnaissance vocale selon la revendication 2, dans lequel ladite réorganisation est faite en parcourant la syntaxe antérieure en fonction des mots de ladite phrase et formation au fil de ce parcours du chemin spécifique à ladite phrase.
- 35 4- Procédé de reconnaissance vocale selon l'une des revendications précédentes, caractérisé en ce que la recherche d'une nouvelle phrase est réitérée de façon systématique pour anticiper la correction.

5- Procédé de reconnaissance vocale selon la revendication 4, caractérisé en ce que chaque nouvelle phrase reconnue est proposée au locuteur sur sa demande.

6- Procédé de reconnaissance vocale selon l'une des 5 revendications 4 ou 5, caractérisé en ce que la recherche d'une nouvelle phrase est stoppée par validation d'une phrase reconnue par le locuteur.

7- Procédé de reconnaissance vocale selon l'une des revendications précédentes, caractérisé en ce que l'étape (13) de traitement comprend :

- 10 - une étape de numérisation et de découpage en une suite de trames temporelles dudit signal acoustique,
 - une phase de paramétrisation de trames temporelles contenant de la parole de manière à obtenir, par trame, un vecteur de paramètres dans le domaine fréquentiel, l'ensemble 15 de ces vecteurs de paramètres formant ledit signal sous sa forme compressée.

8- Procédé de reconnaissance vocale selon la revendication 7, caractérisé en ce que la reconnaissance de forme fait appel à un algorithme de type DTW.

20 9- Procédé de reconnaissance vocale selon la revendication 7, caractérisé en ce que la reconnaissance de forme fait appel à un algorithme de type HMM.